# Transdisciplinary perspectives on validity: bridging the gap between design and implementation for technology-enhanced learning systems

Haastrecht, M.A.N. van

# 3

## RESPITE FOR SMES: A SYSTEMATIC REVIEW

*The contents of this chapter are based on: van Haastrecht, Yigit Ozkan, et al. (2021). Respite for SMEs: A systematic review of socio-technical cybersecurity metrics. Applied Sciences. For the full version including appendices, please consult the original article.*

Cybersecurity threats are on the rise, and small- and medium-sized enterprises (SMEs) struggle to cope with these developments. To combat threats, SMEs must first be willing and able to assess their cybersecurity posture. Cybersecurity risk assessment, generally performed with the help of metrics, provides the basis for an adequate defence. Significant challenges remain, however, especially in the complex socio-technical setting of SMEs. Seemingly basic questions, such as how to aggregate metrics and ensure solution adaptability, are still open to debate. Aggregation and adaptability are vital topics to SMEs, as they require the assimilation of metrics into actionable advice adapted to their situation and needs. To address these issues, we systematically review socio-technical cybersecurity metric research in this chapter. We analyse aggregation and adaptability considerations and investigate how current findings apply to the SME situation. To ensure that we provide valuable insights to researchers and practitioners, we integrate our results in a novel socio-technical cybersecurity framework geared towards the needs of SMEs. Our framework allowed us to determine a glaring need for intuitive, threat-based cybersecurity risk assessment approaches for the least digitally mature SMEs. In future, we hope our framework will help to offer SMEs some deserved respite by guiding the design of suitable cybersecurity assessment solutions.

## 3.1 INTRODUCTION

In recent times, we have seen a surge in cyber threats that businesses are struggling to cope with (Bassett et al., 2021). Additionally, the frequency with which cybersecurity incidents occur, and the costs associated with them, are on the rise (Bissell and Lasalle, 2019). Among businesses, small- and medium-sized enterprises (SMEs) are most vulnerable, due to a shortage of cybersecurity knowledge and resources (Heidt et al., 2019). The vulnerable position of SMEs is being exploited, as witnessed by the large proportion of SMEs that experience cyber incidents (Ponemon Institute, 2019).

In SME cybersecurity, the interplay between the social and the technical is essential (Malatji, Von Solms, et al., 2019), which is why SMEs are often studied from a socio-technical systems (STS) perspective (Carías, Arrizabalaga, et al., 2020). The view of STS is that joint consideration of social and technical elements is necessary (Davis et al., 2014). This view has interesting implications in cybersecurity, where humans are generally found to be the weakest link (Gratian et al., 2018; Shojaifar, Fricker, and Gwerder, 2020).

Due to their lack of resources (Heidt et al., 2019) and the complex socio-technical setting they operate in, SMEs struggle to address their cybersecurity issues autonomously (Benz and Chatterjee, 2020). Before SMEs can begin to improve their cybersecurity posture, it is vital they first assess their current situation (Jaquith, 2007). Assessment of cybersecurity posture is achieved by measuring SME cybersecurity properties, which result in cybersecurity metrics. Regardless of whether measurement results are deemed relevant by the SME, the knowledge gained by those involved in the measurement process is of value (Slayton, 2015). This observation touches once more on the socio-technical nature of the problem, where furthering human knowledge and improving the technical cybersecurity posture of an SME go hand-in-hand.

Cybersecurity assessment generally requires the aid of cybersecurity experts; personnel that SMEs typically do not have (Benz and Chatterjee, 2020; Shojaifar, Fricker, and Gwerder, 2020). A solution to this issue is to automate the cybersecurity assessment process where possible (Shojaifar, Fricker, and Gwerder, 2020). Although automation is a promising approach, the diverse nature of the SME landscape is often ignored (European DIGITAL SME Alliance, 2020; Yigit Ozkan, Spruit, et al., 2019), whereas we know from earlier research that it is vital for SMEs to have solutions adapted to their context and needs (Cholez and Girard, 2014; Mijnhardt et al., 2016).

Another issue is that cybersecurity assessment approaches aimed at SMEs are still scarce (Carías, Arrizabalaga, et al., 2020), explaining why it is not uncommon to see results from other cybersecurity focus areas being applied to the SME setting (Benz and Chatterjee, 2020). Systematic literature reviews are a logical approach to gather knowledge from one focus area, summarise it, and make it available for use in other focus areas.

Systematic reviews that address both the social and technical sides of cybersecurity, already exist (J.-H. Cho et al., 2019; Pendleton et al., 2016). These reviews identified a need for adaptable solutions (J.-H. Cho et al., 2019), which we have seen are also craved by SMEs. Additionally, these papers stress the need for more clarity on how to aggregate security metrics (J.-H. Cho et al., 2019; Pendleton et al., 2016). Given the lack of resources available at SMEs, aggregating information into understandable insights is a requirement for a usable solution (Shojaifar, Fricker, and Gwerder, 2020).

The issue with these systematic reviews is that they offer adaptability and aggregation as areas for future research, rather than addressing the topics head-on. Additionally, they do not provide actionable insights for SMEs since this is not their target audience.

In short, we can conclude that SMEs need (semi-)automated cybersecurity assessment approaches that address their needs for adaptability and aggregation of information. A systematic review offers the potential to gather and summarise such information, providing guidelines for designing usable solutions for SMEs. This motivates the need for a systematic review of cybersecurity metric research, where both the social and technical sides of the puzzle are acknowledged. This is exactly our aim in this chapter, as we try to answer the following research questions:

- **RQ1**: How are cybersecurity metrics aggregated in socio-technical cybersecurity measurement solutions?
- **RQ2**: How do aggregation strategies differ in cybersecurity measurement solutions relevant to SMEs and all other solutions?
  - **RQ2.1**: What are the reasons for these differences?
  - **RQ2.2**: Which aggregation strategies can be used in SME cybersecurity measurement solutions, but currently are not?
- **RQ3**: How do cybersecurity measurement solutions deal with the need for adaptability?

In Section 3.2, we cover related work from several different perspectives to provide a basis for our systematic review. Our systematic review methodology is detailed in Section 3.3, after which we present our results in Section 3.4.

To ensure that the insights we gain on aggregation and adaptability are captured in an actionable form, we incorporate them in a novel socio-technical cybersecurity framework geared towards SME needs. Our framework, introduced in Section 3.5, integrates our systematic review results with existing knowledge to arrive at concise guidelines for what can be expected of various SME categories.

Section 3.6 focuses on outlining the answers to our research questions, as well as covering limitations and threats to validity. Finally, we conclude in Section 3.7, additionally outlining potentially fruitful areas for future research.

## 3.2 RELATED WORK

Before covering work relating to our socio-technical cybersecurity metric setting, we should be clear on our definition of what constitutes a cybersecurity metric. We make use of the definition of a cyber-system as specified in Refsdal et al. (2015): "A cyber-system is a system that makes use of a cyberspace." Refsdal et al. (2015) define cyberspace as "a collection of interconnected computerized networks, including services, computer systems, embedded processors, and controllers, as well as information in storage or transit." There is no standard definition of what constitutes a (cyber)security metric (Pendleton et al., 2016). Borrowing ingredients from earlier definitions, we define a cybersecurity metric to be *any value resulting from the measurement of security-related properties of a cyber-system* (Böhme and Freiling, 2008; Pendleton et al., 2016; Refsdal et al., 2015).

### 3.2.1 *Socio-technical cybersecurity*

Humans are often considered the weakest link in cybersecurity (Martens et al., 2019). It is vital to recognise the interaction of the social and technical sides of cyber-systems when modelling and measuring cybersecurity, which is why the field of STS has played such an important role in cybersecurity metric research (Gollmann et al., 2015). STS research has uncovered the dangers of considering social and technical elements separately (Selbst et al., 2019) and has offered insight into how to avoid these dangers (Davis et al., 2014).

Recognition of the human factor in cybersecurity goes beyond simply including static human actors. This is where behavioural theories such as Protection Motivation Theory (PMT) and Self-Determination Theory (SDT) come in (Menard et al., 2017; Padayachee, 2012). PMT reserves a prominent role for extrinsic motivators and threat appraisal (Herath and Rao, 2009). SDT includes extrinsic motivation as a central concept but often focuses on moving from extrinsic to increasingly internalised motivation (Padayachee, 2012). In the context of SMEs, intrinsic motivation to improve cybersecurity is often hard to find. However, there are solutions to this problem. Committing to improving cybersecurity in an organisation can motivate employees (Padayachee, 2012). From the STS perspective, it is common to distinguish between metrics that include the real-life threat environment and those that do not (Gollmann et al., 2015). Threat perception lies at the core of PMT and is important in security applications using SDT (Menard et al., 2017). Another solution to promote motivation among SME employees would therefore be to incorporate the real-life threat environment in our cybersecurity metrics. Later in this chapter, in Section 3.4, we describe whether this is indeed something we observe in current research.

We will address the social dimension using the ADKAR model of Hiatt (2006). This model, originating from change management, considers five

Table 3.1: Existing cybersecurity metric (systematic) reviews. The research focus area is shown, with 'generic' indicating research without a specific focus area. We consider social factors to be evaluated when the review covers socio-technical cybersecurity metrics.

| RESEARCH | YEAR | FOCUS AREA | SOCIAL FACTORS |
|---|---|---|---|
| **Current chapter** | **2021** | **Generic** | ✓ |
| Verendel (2009) | 2009 | Generic | × |
| Rudolph and Schwarz (2012) | 2012 | Generic | × |
| Pendleton et al. (2016) | 2016 | Generic | ✓ |
| J.-H. Cho et al. (2019) | 2019 | Generic | ✓ |
| Husák, Komárková, et al. (2019) | 2019 | Attack Prediction | ✓ |
| Iannacone and Bridges (2020) | 2020 | Cyber Defense | × |
| Kordy et al. (2014) | 2014 | Directed Acyclic Graphs | × |
| Cadena et al. (2020) | 2020 | Incident Management | ✓ |
| Knowles et al. (2015) | 2015 | Industrial Control Systems | ✓ |
| Asghar et al. (2019) | 2019 | Industrial Control Systems | ✓ |
| Eckhart et al. (2019) | 2019 | Industrial Control Systems | × |
| Jing et al. (2019) | 2019 | Internet Security | × |
| Sengupta et al. (2020) | 2020 | Moving Target Defense | × |
| Liang and Xiao (2013) | 2013 | Network Security | × |
| Ramos et al. (2017) | 2017 | Network Security | ✓ |
| Cherdantseva et al. (2016) | 2016 | SCADA Systems | ✓ |
| Morrison et al. (2018) | 2018 | Software Security | × |
| W. He et al. (2019) | 2019 | Unknown Vulnerabilities | × |
| Xie et al. (2019) | 2019 | Wireless Networks | × |

phases in managing the personal side of change: awareness, desire, knowledge, ability, and reinforcement. ADKAR has previously been applied in assessing information security culture within organisations (Da Veiga, 2018). We apply ADKAR as a means to classify the socio-technical cybersecurity metrics we encounter. We define a socio-technical cybersecurity metric to be *a cybersecurity metric that requires measuring the outcome(s) of the actions of at least one (simulated) human actor.* We do not address the technical dimension explicitly in this definition, as the technical dimension is implicit in the term 'cybersecurity.' We hypothesise that all socio-technical cybersecurity metrics can be linked to one or more of the ADKAR categories.

### 3.2.2 *Cybersecurity metric reviews*

Systematic reviews are common in cybersecurity metric research. However, as Table 3.1 shows, they are often narrow in scope. Either the focus area is narrow, or the research does not consider social factors. The papers that do cover both social and technical factors, often do so passingly, and without covering the intricacies and implications of socio-technical interactions.

Some exceptions are comprehensive and cover both social and technical factors (J.-H. Cho et al., 2019; Pendleton et al., 2016). Interestingly, exactly

these papers outline that future research should focus on "how to aggregate and to what extent to aggregate" (Pendleton et al., 2016). Additionally, they stress the importance of adaptability, meaning by this "the state of being able to change to work or fit better" (J.-H. Cho et al., 2019). This need for adaptability has been confirmed by experience from practice (Ray et al., 2020).

We address the acknowledged challenges of aggregation and adaptability head-on in our systematic review, ensuring that our approach is both distinct from earlier work and provides a meaningful contribution to the field. Furthermore, we employ a novel systematic review approach (as outlined in Section 3.3) and target our analysis to aid SMEs, a group with specific needs often not considered in earlier work.

### 3.2.3 *Aggregation*

In cybersecurity metric research, aggregation strategies vary, although the importance of proper aggregation is widely recognised (J.-H. Cho et al., 2019; Pendleton et al., 2016). To discuss different aggregation strategies, we define a mathematical context with an aggregation strategy $S : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$, where $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers. We define metric value variables $x_i$, corresponding to metrics $i = 1, \ldots, n$. The metric values are assumed to be non-negative: $x_i \in \mathbb{R}_{\geq 0}\ \forall i$. We assume that for each metric, a higher metric value corresponds to lower security, without loss of generality. A negative relationship between a metric and security is common in the security literature, as it is often the lack of security, or risk, which is being measured.

A desirable property of a strategy $S$ is that it is responsive to changes in metric values. This is captured by the property of injectivity, where we consider a strategy $S$ to be injective when for $a, b \in \mathbb{R}_{\geq 0}$, $a \neq b$, $S(a, x_1, x_2, \ldots, x_n) \neq S(b, x_1, x_2, \ldots, x_n)$. Injectivity implies that a change in a metric value will always result in a change of the aggregate, provided all else remains constant. A stronger requirement would be strict monotonicity of the strategy $S$. Although this property could be desirable in the cybersecurity context, we only consider the less strict injectivity in this chapter.

A common property of averages, which constitute a specific branch of aggregation, is idempotence. A strategy $S$ is idempotent, when for $a \in \mathbb{R}_{\geq 0}$, $S(a, a, \ldots, a) = a$. When an aggregation strategy $S$ is both injective and idempotent, the result of the aggregation always lies between the minimum and the maximum values of all metrics. Both injectivity and idempotence capture what we would intuitively expect of an aggregation strategy, as these are properties satisfied by the Pythagorean means. In this sense, these are desirable properties in the context of SMEs, where cybersecurity knowledge is often lacking. To still allow employees to feel competence and relatedness (Menard et al., 2017) in the complex cybersecurity setting, we should at least use an aggregation strategy they understand.

Three additional properties are important in the security context. The possibility to prioritise certain metrics over others is desirable (Lippmann and Riordan, 2016). Formally, we consider a strategy to allow for prioritisation when for any $a, b > 0$, $a \neq b$, there exists a pair $i, j$ with $i \neq j$, such that $S(x_1, \ldots, x_i = a, \ldots, x_j = b, \ldots, x_n) \neq S(x_1, \ldots, x_i = b, \ldots, x_j = a, \ldots, x_n)$.

Strategies should also be able to accommodate dependencies between security metrics. However, it is complicated to include metric dependencies, with some seeing it as "the most challenging task" in aggregation (J.-H. Cho et al., 2019). For strategies in the set $\mathbb{D}$ of strategies that satisfy the necessary differentiability properties, we define a strategy $S$ to allow for dependencies, when there exist distinct metrics $i$, $j$, and $k$ such that:

$$\frac{\partial^2 S}{\partial x_i \partial x_j} \not\propto \frac{\partial^2 S}{\partial x_i \partial x_k}. \tag{3.1}$$

Equation 3.1 captures the idea that a strategy $S$ allows for dependencies among metrics when it allows for relationships among metrics that are not proportional to other relationships. For aggregation strategies $S \notin \mathbb{D}$, we employ the same verbal definition. Care should be taken to adjust the criterion of Equation 3.1 appropriately where it cannot be applied directly for the strategy $S$.

A last core principle in security is that systems are only as secure as their weakest link (N. Ferguson and Schneier, 2003). Assuming that we have at least two distinct values among our metrics, there exists a minimum value $x_{min}$ and a maximum value $x_{max}$. Since we assume metrics relate negatively to security, $x_{max}$ corresponds to the weakest link. A strategy $S$ satisfies the weakest link principle if for any $a > 0$, $S(x_{min} + a, \ldots, x_{max}) \leq S(x_{min}, \ldots, x_{max} + a)$, and there exists an $\alpha > 0$, such that $S(x_{min} + \alpha, \ldots, x_{max}) < S(x_{min}, \ldots, x_{max} + \alpha)$. Thus, weakening the weakest link has more impact than weakening the strongest link with an equal amount.

The most common aggregation strategy employed in the literature is the weighted linear combination (WLC), which can be defined as:

$$S_{WLC}(\mathbf{x}) = a + \frac{\sum_{i=1}^{n} w_i \cdot x_i}{b},\ a \geq 0,\ b > 0,\ w_i > 0\ \forall i. \tag{3.2}$$

WLC contains the special cases of the weighted sum ($a = 0,\ b = 1$), the weighted average ($a = 0,\ b = \sum w_i$), and the arithmetic mean ($a = 0,\ b = n,\ w_i = 1\ \forall i$). WLC strategies are injective, idempotent, and allow for prioritisation through weighting. However, these strategies do not allow for dependencies and do not satisfy the weakest link principle.

A related set of strategies are the weighted product (WP) strategies:

$$S_{WP}(\mathbf{x}) = a + b \cdot \prod_{i=1}^{n} x_i^{w_i},\ a \geq 0,\ b > 0,\ w_i \in (0, 1]\ \forall i. \tag{3.3}$$

Among the WP strategies are the simple product ($a = 0,\ b = 1,\ w_i = 1\ \forall i$) and the geometric mean ($a = 0,\ b = 1,\ w_i = \frac{1}{n}\ \forall i$). WP strategies satisfy the same properties as WLC strategies, except for the idempotence property which these strategies do not satisfy.

Using the weighted maximum (WM) - $S_{WM}(\mathbf{x}) = max\{w_1 \cdot x_1, \ldots, w_n \cdot x_n\},\ w_i > 0\ \forall i$ - metric value as the aggregated value is uncommon in most disciplines, since this strategy is not injective. However, it is used in the security field (Lippmann, Riordan, et al., 2012), and is in fact an extreme case of satisfying the weakest link principle. WM allows for prioritisation, although the basic maximum function does not.

The complementary product is another aggregation strategy that is uncommon outside of the security field (Lippmann, Riordan, et al., 2012). Let $\hat{x}_i$, for $i = 1, 2, \ldots, n$, denote the metric value normalised to $[0, 1)$. Let $w_i$ be the weight of metric $i$ for $i = 1, 2, \ldots, n$. We define the weighted complementary product (WCP) class as:

$$S_{WCP}(\mathbf{x}) = a \cdot \left(1 - \prod_{i=1}^{n}(1 - \hat{x}_i)^{w_i}\right),\ a > 0,\ w_i \in (0, 1]\ \forall i. \tag{3.4}$$

The regular complementary product is achieved with $a = 1$ and $w_i = 1\ \forall i$. WCP strategies are injective and can satisfy the prioritisation and weakest link principles, depending on the values of $w_i$.

None of the strategies considered so far consider dependency. Bayesian networks (BN) are probabilistic graphical models, often of a causal nature, that are commonly applied in the security field (Kordy et al., 2014). In BN aggregation strategies, the metric values $x_i$ are assumed to originate from discrete, bounded random variables $X_i$, corresponding to the metrics $i = 1, \ldots, n$. The conditional dependencies between the random variables, and with a potential unobserved variable $Y$, are made explicit. This allows us to infer the probabilities of different values of $Y$, based on the metric values $x_i$. BN strategies are injective, but not idempotent. Although prioritisation is generally not a goal within these strategies, the prioritisation property will usually be satisfied. BN strategies accommodate dependencies by their nature, but will mostly not satisfy the weakest link principle.

The strategy classes presented in Table 3.2 are not exhaustive but do cover the large majority of all aggregation strategies employed, as we show in Section 3.4. Two examples of other possibilities are the use of analytic network process (ANP) techniques (Brožová et al., 2016; Lo and W.-J. Chen, 2012), which relate to the deterministic equivalent of Bayesian networks, and the analysis of game-theoretic equilibria (Rass et al., 2017). What is common to all strategies, is that none satisfy all criteria of Table 3.2, where we should additionally note that strategies within the classes of weighted maximum and weighted complementary product cannot satisfy the prioritisation and weakest link properties at the same time.

Table 3.2: Various classes of metric aggregation strategies, and important security-related properties their strategies can possess.

| AGGREGATION | INJECTIVE | IDEMPOTENT | PRIORITISATION | DEPENDENCE | WEAKEST LINK |
|---|---|---|---|---|---|
| Weighted linear combination | ✓ | ✓ | ✓ | × | × |
| Weighted product | ✓ | × | ✓ | × | × |
| Weighted maximum | × | ✓ | ✓ | × | ✓ |
| Weighted complementary product | ✓ | × | ✓ | × | ✓ |
| Bayesian network | ✓ | × | ✓ | ✓ | × |

### 3.2.4 *Adaptability*

Adaptability is crucial to any cybersecurity solution (Evesti and Ovaska, 2013). Especially when measuring cybersecurity, a rigid solution that does not adapt to a changing environment or a new use case is far from optimal (Baars et al., 2016). It is not surprising to see, then, that adaptability is a key focus of many studies (de las Cuevas et al., 2015; Yigit Ozkan, Spruit, et al., 2019), although operationalisation of adaptability is still a challenge (Evesti and Ovaska, 2013).

We consider adaptability to be "the state of being able to change to work or fit better" (J.-H. Cho et al., 2019). This definition outlines two important dimensions of adaptability. Firstly, a solution is considered adaptable if it can change to work better. There are several reasons why a cybersecurity metric solution may not be functioning as it should. This can relate to problems with the metrics themselves, such as missing or dirty data (W. Kim et al., 2003). It can also relate to a changing security landscape, that invalidates an existing model. This phenomenon is known as concept drift (Widmer and Kubat, 1996). Secondly, a solution is considered adaptable if it can change to fit better. Generally, cybersecurity solutions in research are made to fit their use case. We can determine their adaptability in the 'fitting' dimension by determining how easily the solution can be deployed at other (similar) use cases.

Adaptability is significant in the SME context. The SME landscape is diverse (European DIGITAL SME Alliance, 2020), and SMEs often lack the knowledge and expertise to perform extensive adaptations independently (Shojaifar, Fricker, and Gwerder, 2020). In Section 3.6, we assimilate observations from earlier research and our results of Section 3.4 to provide suggestions for improving solution adaptability.

## 3.3 SYSTEMATIC REVIEW METHODOLOGY

We performed a systematic literature review to address our research questions. To ensure broad coverage of the cybersecurity metrics field, we employed a novel Systematic Review Methodology Blending Active Learning and Snowballing (SYMBALS, (van Haastrecht, Sarhan, Yigit Ozkan, et al., 2021)), which

combines existing methods into a swift and accessible methodology, while following authoritative systematic review guidelines (Kitchenham and Charters, 2007; Liberati et al., 2009; Moher et al., 2015).

Active learning is one of the cornerstones of the SYMBALS approach. Active learning is commonly applied in the title and abstract screening phase of systematic reviews, where researchers start with a large set of papers and prefer to not screen them all manually (van de Schoot et al., 2021). Active learning is uniquely suited to this task, as this machine learning method selects the ideal data points for an algorithm to learn from.

SYMBALS complements active learning with backward snowballing. From a set of included papers, a researcher can find additional relevant papers by consulting references (backward snowballing) and citations (forward snowballing) (Wohlin, 2014). Snowballing has proven to be a valuable addition to systematic reviews, even when reviews already include an extensive database search (Mourão, Pimentel, et al., 2020). Backward snowballing is especially useful in uncovering older relevant research. Forward snowballing is not employed within SYMBALS, based on the observation that databases generally have excellent coverage of recent peer-reviewed research.

After the development and evaluation of a systematic review protocol for this research, we commenced with the database search step of SYMBALS. We retrieved research from abstract databases (Scopus, Web of Science) and full-text databases (ACM Digital Library, IEEE Xplore, PubMed Central).

The Scopus API was used to retrieve an initial set of relevant research. Results from other sources were then successively added to this set. The order in which sources were consulted can be surmised from Table 3.3. The Python Scopus API wrapper 'pybliometrics' (Rose and Kitchin, 2019) was used to retrieve all research available through the Scopus API, that satisfied the query:

```
AUTHKEY((security* OR cyber*)
AND (assess* OR evaluat* OR measur* OR metric* OR model* OR risk*
    OR scor*))
AND LANGUAGE(english) AND DOCTYPE(ar OR bk OR ch OR cp OR cr OR re
    )
```

The 'AUTHKEY' field corresponds to the keywords that authors provided for a paper. Our search query is intentionally broad, as the SYMBALS methodology allows us to deal with larger quantities of research, and we aim to exclude as little relevant research as possible at this stage. We did choose to only include English language research and document types where extensive and verifiable motivations for findings can be reported.

Table 3.3 summarises the query results. ACM Digital Library and IEEE Xplore limit the number of accessible papers to 2,000. This means only the 2,000 most relevant papers from these sources could be considered. Moreover, IEEE Xplore only allows the use of 6 wildcards in the search query. We removed the 'security' and 'cyber' wildcards for the IEEE Xplore search to comply with this limitation. Any research without an abstract was excluded,

Table 3.3: Statistics regarding the different databases used in the search procedure.

| SOURCE | RESULTS | UNIQUE |
|---|---|---|
| Scopus | 21,964 | 21,964 |
| Web of Science | 7,889 | 1,782 |
| ACM Digital Library | 2,000 | 660 |
| IEEE Xplore | 2,000 | 1,256 |
| PubMed Central | 660 | 111 |
| **Total** | **34,513** | **25,773** |

as this is vital to the active learning phase of SYMBALS. This led to a small set of exclusions from the PubMed Central database. Duplicate removal was performed based on the research title, although we found that this process was not perfect, due to different character sets being accepted in different databases.

Altogether, our dataset resulting from database search comprised 25,773 papers. This exemplifies the broad scope of our research, as the largest initial set of papers from the reviews in Table 3.1 comprised 4,818 papers (Morrison et al., 2018).

The set of 25,773 papers is too large to perform data extraction directly. This is where the active learning phase of SYMBALS comes in. We chose to use ASReview in this phase, a tool that offers active learning capabilities for systematic reviews, specifically for the title and abstract screening step (van de Schoot et al., 2021). Many other active learning tools exist that are worth considering (Harrison et al., 2020). However, we found ASReview effective and easy to use, and additionally value the commitment its developers have made to open science. This shows in, among other things, the codebase that they made available open-source.

In the ASReview process, as well as in the later review phases, we made use of the following inclusion and exclusion criteria:

- Inclusion criteria:
  - I1: The research concerns cybersecurity metrics and discusses how these metrics can be used to assess the security of a (hypothetical) cyber-system.
  - I2: The research is a review of relevant papers.
- Exclusion criteria:
  - E1: The research does not concern cyber-systems.
  - E2: The research does not describe a concrete path towards calculating cybersecurity metrics (only applied if I2 is not applicable).
  - E3: The research has been retracted.
  - E4: There is a more relevant version of the research that is included.

- E5: The research was automatically excluded due to its assessed irrelevance by the ASReview tool.
- E6: The research does not satisfy the database query criteria on language and document type.
- E7: No full-text version of the research can be obtained.
- E8: The research is of insufficient quality.
- E9: The research does not contain at least one socio-technical cybersecurity metric.

Exclusion criterion E8 relates to the quality assessment phase of SYMBALS, which is explained below. Criterion E9 requires the consideration of the full text to be determined, as abstracts do not contain enough information to make a decision regarding this intricate topic (Brereton et al., 2007). Thus, neither of these criteria were applied during title and abstract screening.

ASReview requires users to specify prior relevant and irrelevant papers to train its algorithm. We used five papers as initial indications of relevance to ASReview (Allodi and Massacci, 2017; J.-H. Cho et al., 2019; Noel and Jajodia, 2014; Spruit and Röling, 2014; Stolfo et al., 2011). These papers were chosen since they cover diverse topics, were written by different authors at different times and were published in different journals and conferences. ASReview additionally provides the option to label a certain number of random papers before proceeding, assuming that a significant proportion of these papers will be irrelevant. This provides the algorithm with a balance of relevant and irrelevant papers for training. We labelled 5 random papers, giving us a total training set of 10 papers.

The ASReview tool then presents the paper whose classification it deems most informative to learn from. The tool quickly learns to distinguish between relevant and irrelevant papers. By presenting the researcher mostly relevant papers, the process of discovering relevant papers is accelerated.

Although ASReview offers several classifier options, we employed the default Naïve Bayes classifier using term frequency-inverse document frequency (TF-IDF) feature extraction and certainty-based sampling. The default settings have been shown to produce consistently good results and are additionally commonly available in other active learning tools (van de Schoot et al., 2021). Thus, our decision to use the default settings can be motivated both from a performance and a reproducibility standpoint.

At some point in the active learning process, mostly irrelevant research remains. To reduce the time spent on assessing irrelevant research, a stopping criterion is used (van de Schoot et al., 2021). We stop evaluating research when the last 20 reviewed papers were considered irrelevant, although more sophisticated stopping criteria exist that are worth considering (Cormack and Grossman, 2016). All research that was not evaluated at this stage, was

Table 3.4: The quality criteria applied to 60 papers during the quality assessment phase. Possible responses were strongly disagree (SD), disagree (D), neutral (N), agree (A), or strongly agree (SA).

| ASPECT | CRITERION | SD | D | N | A | SA |
|---|---|---|---|---|---|---|
| Reporting | There is a clear statement of the research aims. | 0 | 4 | 7 | 28 | 21 |
| | There is an adequate description of the research context. | 0 | 6 | 11 | 17 | 26 |
| | The paper is based on research. | 0 | 3 | 3 | 16 | 38 |
| Rigour | Metrics used in the study are clearly defined. | 0 | 10 | 19 | 16 | 15 |
| | Metrics are adequately measured and validated. | 1 | 24 | 22 | 8 | 5 |
| | The data analysis is sufficiently rigorous. | 0 | 21 | 17 | 14 | 8 |
| Credibility | Findings are clearly stated and related to research aims. | 0 | 8 | 19 | 25 | 8 |
| | Limitations and threats to validity are adequately discussed. | 30 | 18 | 8 | 2 | 2 |
| Relevance | The study is of value to research and/or practice. | 0 | 9 | 12 | 28 | 11 |

excluded based on exclusion criterion E5. As Figure 3.1 shows, 1,644 papers remained after the active learning phase.


Figure 3.1: Visualisation of the SYMBALS steps as applied in our cybersecurity metric systematic review.

We then proceeded with the backward snowballing phase of SYMBALS. We followed the ASReview evaluation order in our backward snowballing procedure. We concluded backward snowballing once 10 consecutive papers contained no new references satisfying the inclusion criteria. As can be seen in Figure 3.1, 1,796 papers were contained in our inclusion set after the completion of this phase.

SYMBALS specifies quality assessment as an optional step, but given the large number of papers remaining, assessing quality was deemed necessary. Table 3.4 outlines the quality criteria that were applied. Commonly used research quality criteria were adapted for use with a Likert scale (Y. Zhou et al., 2015). Statements could be responded to with strongly disagree, disagree, neutral, agree, or strongly agree. Instead of applying these criteria to all 1,796 inclusions, the two researchers involved in quality assessment evaluated 40 papers, with 20 papers being evaluated by both researchers.

A simple, yet effective, solution to extrapolate these results is to train a binary decision tree on basic research characteristics, to create a model that can distinguish research of sufficient quality from research of insufficient quality. The five Likert scale responses were assigned scores of 0 (strongly disagree), 0.25 (disagree), 0.5 (neutral), 0.75 (agree), and 1 (strongly agree). Summing the

quality criteria scores, each paper received a score between 0 and 9. To make the problem a binary decision problem, we labelled papers with a score of at least 6 as having sufficient quality. The height of this threshold determines how strict the eventual model will be.

Next, we split our set of 60 evaluated papers into a training set of 48 papers (80%) and a test set of 12 papers (20%). To be able to train a model on this set, we need explanatory variables which explain the quality scores obtained by the papers. We opted to use three features: years since publication, citation count, and the number of pages. The maximum depth of the binary decision tree was set to 3, meaning at most 3 binary splits are performed before classifying a paper as having sufficient or insufficient quality. The model was trained on the 48 training papers and evaluated on the 12 test papers. Despite - or perhaps because of - the model's simplicity, 11 of the 12 test papers were labelled correctly. The only incorrect labelling occurred in an edge case with a quality score of 6. Similar results were obtained in replications with different random seeds. Figure 3.1 shows that 516 papers remained after applying the binary decision tree to our complete inclusion set.

Finally, we applied exclusion criterion E9 using a manual screening process, to filter out the papers that do not consider the social side of cybersecurity, as defined in Section 3.5. Figure 3.1 shows that in total 60 papers were included after our filtering step.

## 3.4 RESULTS

In this section, we focus on descriptive analysis of aggregate results. In Sections 3.5 and 3.6, we will dive deeper, to interpret and contextualise the results.

Figure 3.2 depicts the relative prevalence of each of the five ADKAR factors over the years. Since 2010, awareness and reinforcement together constituted over half of the ADKAR considerations. Desire is the element that receives the least attention in research. Table 3.5 lists the related concepts which we encountered and mapped to each of the ADKAR terms.

Part of the reason for the prevalence of reinforcement research is that cybersecurity training and education belong to this ADKAR element. Researchers feel that organisational reinforcement is an important aspect of the social side of cybersecurity. At the same time, reinforcement can be easier to measure than other factors, which may offer a partial explanation for its prevalence. For example, many researchers choose to include a metric of cybersecurity awareness training (reinforcement), rather than of cybersecurity awareness itself (awareness).

Various security concepts were assessed in our inclusions, as shown in Table 3.6. Some researchers choose to measure security itself (Bhilare et al., 2008; You et al., 2015), but this approach is too general for most. Risk was assessed in two-thirds of all papers. This is interesting, as risk can be seen as having

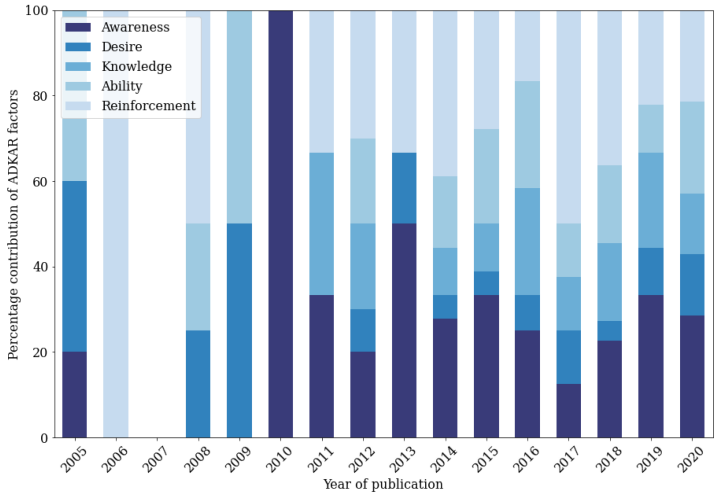


Figure 3.2: The consideration of the five ADKAR factors over the years, based on the 60 inclusions of our systematic review.

Table 3.5: The ADKAR factors and the related concepts we encountered which were associated to each factor.

| ADKAR | ABBREVIATION | RELATED CONCEPTS |
|---|---|---|
| Awareness | AW | Consciousness |
| Desire | DE | Motivation, loyalty, attendance |
| Knowledge | KN | Understanding |
| Ability | AB | Behaviour, capability, capacity, experience, skill |
| Reinforcement | RE | Culture, education, evaluation, policy, training |

Table 3.6: The various security assessment concepts discussed in research, with an indication of the ADKAR elements covered and the aggregation strategies employed. Each paper should consider at least one ADKAR element. A paper may not aggregate at all, but could also employ several aggregation strategies. Reviews were not labelled with a specific assessment concept.

| | | ADKAR elements | | | | | Aggregation strategy classes | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CONCEPT | TOTAL | AW | DE | KN | AB | RE | WLC | WP | WM | WCP | BN | NONE |
| Risk | 40 | 24 | 9 | 14 | 19 | 28 | 27 | 10 | 7 | 1 | 4 | 4 |
| Awareness | 5 | 5 | 3 | 4 | 3 | 2 | 3 | 1 | 1 | 0 | 0 | 2 |
| Maturity | 5 | 0 | 0 | 0 | 0 | 5 | 4 | 0 | 1 | 0 | 0 | 0 |
| Resilience | 3 | 3 | 1 | 0 | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 0 |
| Security | 2 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 1 |
| Vulnerability | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

a negative connotation, whereas awareness, maturity, and resilience have positive connotations. This finding conflicts with the general tendency in the security community to favour SDT approaches over the fear- and threat-based approaches more associated with PMT (Menard et al., 2017), especially in the context of organisations (N. Yang et al., 2020).

When analysing the ADKAR factors by assessment concept, the papers assessing security maturity stand out. These papers place a large focus on the organisational reinforcement of security and ignore all other ADKAR factors. This is not a surprising finding. Maturity is generally a concept that requires an assessment of the organisation, rather than the individuals who make up this organisation.

Table 3.6 shows that most papers stick to WLC, WP, and WM as aggregation strategies. It is worth pointing out that not aggregating is a reasonable choice. If it is not necessary for a particular context, it should be avoided, based on our conclusion from Table 3.2 that no aggregation method satisfies all ideal security properties.

Table 3.7 focuses on the actors that were considered from the social viewpoint. Almost all papers focus solely on the defender. It is interesting to see that the desire and ability factors of ADKAR are much more prominent in research including the attacker. We would expect to see more focus from research on desire, and the related concept of motivation, based on the important role that motivation and internalisation play in SDT and PMT (Padayachee, 2012). Desire and motivation are not easily measurable concepts, but metrics such as 'attendance at security sessions' can serve as useful proxies here (Manifavas et al., 2014).

Nearly all research that considers the attacker perspective considers the real-life threat environment as specified in Gollmann et al. (2015). In papers covering the defender, it is quite common to ignore threats entirely (Y. Shin et al., 2011) or to use a proxy such as the prevalence of vulnerabilities to

Table 3.7: The different social viewpoints considered in our inclusions.

| | | ADKAR | | | | | |
|---|---|---|---|---|---|---|---|
| SOCIAL VIEWPOINT | TOTAL | AW | DE | KN | AB | RE | REAL-LIFE THREAT |
| Defender | 52 | 33 | 7 | 17 | 17 | 37 | 18 |
| Attacker | 5 | 0 | 4 | 1 | 5 | 0 | 5 |
| Both | 3 | 2 | 3 | 1 | 3 | 3 | 2 |

Table 3.8: Different aggregation strategy classes and the situations in which they were employed.

| | Classification | | |
|---|---|---|---|
| AGGREGATION STRATEGY | THEORETICAL | IMPLEMENTATION | REVIEW |
| WLC | 38 | 1 | 3 |
| WP | 11 | 0 | 0 |
| WM | 8 | 1 | 0 |
| WCP | 1 | 0 | 0 |
| BN | 4 | 0 | 0 |
| None | 7 | 2 | 1 |

represent threats (Marconato et al., 2013). This is remarkable given the vital role that threat perception plays in both SDT and PMT (Menard et al., 2017).

Table 3.8 groups research based on the employed aggregation strategy. Inclusions were classified into one of three classes: theoretical, implementation, or review. The research was classified as an implementation if either clear and described actions were taken based on the implemented method, or the model was assessed at more than one point in time. This strict requirement explains why most papers were classed as theoretical.

One immediately notices from Table 3.8 that two of the four implementation papers do not employ an aggregation strategy. As we discussed in Section 3.2.3 and showed in Table 3.2, aggregation should only be carried out if deemed necessary. In half of the implementation research of our inclusions, researchers felt the benefits of aggregation did not outweigh the drawbacks.

We additionally see that most research sticks to WLC and WP strategies, which do not satisfy the weakest link principle and cannot take into account dependencies. Researchers prefer simple and explainable strategies, that are injective or idempotent, over strategies that satisfy more security properties. Out of our 60 inclusions, 10 used fuzzy logic approaches. Although translating qualitative statements to fuzzy numbers differentiates these methods from approaches using crisp numbers, most still use some combination of WLC, WP, and WM to aggregate (for example, (X. Li et al., 2018; Shameli-Sendi, Shajari, et al., 2012; Silva et al., 2014)).

Exceptions are Lo and W.-J. Chen (2012) and Brožová et al. (2016), who use an ANP approach to capture dependencies. Lo and W.-J. Chen (2012), Brožová et al. (2016) and the four papers using a bayesian network approach (Dantu

Table 3.9: ADKAR and aggregation strategy frequencies of enterprise research and other research.

| | | Application area | | |
|---|---|---|---|---|
| PROPERTY | VALUES | ANY ENTERPRISE | M/L ENTERPRISE | OTHER |
| ADKAR | AW | 9 | 6 | 20 |
| | DE | 3 | 1 | 10 |
| | KN | 7 | 2 | 10 |
| | AB | 6 | 3 | 16 |
| | RE | 11 | 13 | 15 |
| Aggregation | WLC | 13 | 7 | 22 |
| | WP | 0 | 3 | 8 |
| | WM | 2 | 2 | 5 |
| | WCP | 0 | 0 | 1 |
| | BN | 0 | 1 | 3 |
| | None | 1 | 4 | 5 |

and Kolan, 2005; Dantu, Kolan, and Cangussu, 2009; N. Feng et al., 2014; Sahinoglu, 2008) are the only papers that consider dependencies between metrics. Interestingly, all of these papers were published in 2016 or earlier. It is not immediately clear what the underlying reason is for the current drought in research considering dependencies, but it is certainly a research area that deserves more attention.

Table 3.9 provides detailed results regarding the research application area. Although more enterprise sizes were considered, we only encountered research applicable to medium- and large-sized enterprises, and research applicable to any enterprise size. As with research focused on maturity modelling, we see a strong focus on the reinforcement factor of ADKAR in enterprise research, especially for larger enterprises.

In research intended to apply to any enterprise, Table 3.9 shows that WLC is by far the most popular aggregation strategy class. The only other strategy class that is used is WM. We believe it is not a coincidence that these are the only aggregation strategy classes that are both injective and idempotent. Strategies with these properties are likely to be more intuitive and easy to understand, as explained in Section 3.2.3. Therefore, it is not surprising that these strategies are proposed in research addressing all enterprise sizes, since especially smaller businesses need to be motivated through approachable solutions.

Regarding adaptability, of the 56 inclusions that were not review papers, 44 do not make any consideration for missing or dirty data. Of the papers that do consider one or both of these issues, the most common strategy is to ignore the associated problems. Out of these 56 papers, 46 are not able to adapt to a security event occurring, mostly since they do not operate in a live setting, but are formulated as periodic assessments. Even then, most authors do not cover this topic, and it is certainly not always clear how the security assessment would be adapted after an incident.

Concept drift and adaptation to other use cases are also often not considered. Just four of our inclusions explicitly consider concept drift and no paper mentions a concrete timeline for when a solution should be updated. Adaptation to other use cases is discussed in 24 of our inclusions. However, the majority of these papers only give a rough outline of how the solution could be adapted. A better practice would be to give concrete guidelines on how to adapt the solution or to immediately analyse several use cases. The former approach was not seen in research, whereas the latter was (for example, (Chan, 2011; M.-K. Chen and S.-C. Wang, 2010; Luh et al., 2020; Proença and Borbinha, 2018)).

## 3.5 SOCIO-TECHNICAL CYBERSECURITY FRAMEWORK FOR SMES

To offer more insight into how we can create effective cybersecurity assessment solutions for SMEs, we position our results and findings in the STS analysis framework of Davis et al. (2014). Figure 3.3 shows the view of STS as consisting of six internal social and technical aspects, within an external environment. We rename the 'Buildings/Infrastructure' aspect of Davis et al. (2014) to 'Assets.' This ensures that our view is better aligned with standard terminology in cybersecurity literature. Based on the importance of policies in socio-technical cybersecurity frameworks (Malatji, Von Solms, et al., 2019), we explicitly include policies in the 'Processes/Procedures' aspect of Davis et al. (2014) and rename this aspect to 'Processes.'

The socio-technical system we study is the SME, in the context of cybersecurity. However, the complete set of SMEs is too diverse to consider this group as a single collective. This is why the European DIGITAL SME Alliance proposes to use four SME categories, based on the different roles SMEs can play in the digital ecosystem: start-ups, digitally dependent SMEs, digitally based SMEs, and digital enablers (European DIGITAL SME Alliance, 2020). The European DIGITAL SME Alliance specifies these categories in the context of cybersecurity standardisation, which is intricately related to our cybersecurity assessment setting, making it a suitable classification.

The European DIGITAL SME Alliance defines start-ups as SMEs where "security has a low priority." They "typically neglect (or are not aware of) requirements" for running a secure business. Digitally dependent SMEs are companies that depend on digital solutions (as end users) to run their business. Digitally based SMEs "highly depend on digital solutions for their business model," and, finally, digital enablers are SMEs that develop and provide digital solutions (European DIGITAL SME Alliance, 2020).

Table 3.10 introduces our framework, which synthesises the SME categories of the European DIGITAL SME Alliance (2020) with the STS aspects of Davis et al. (2014). Each SME category has different cybersecurity goals based on their different roles in the digital ecosystem. In Table 3.10, the SME categories are ordered from least to most mature regarding cybersecurity. We expect the

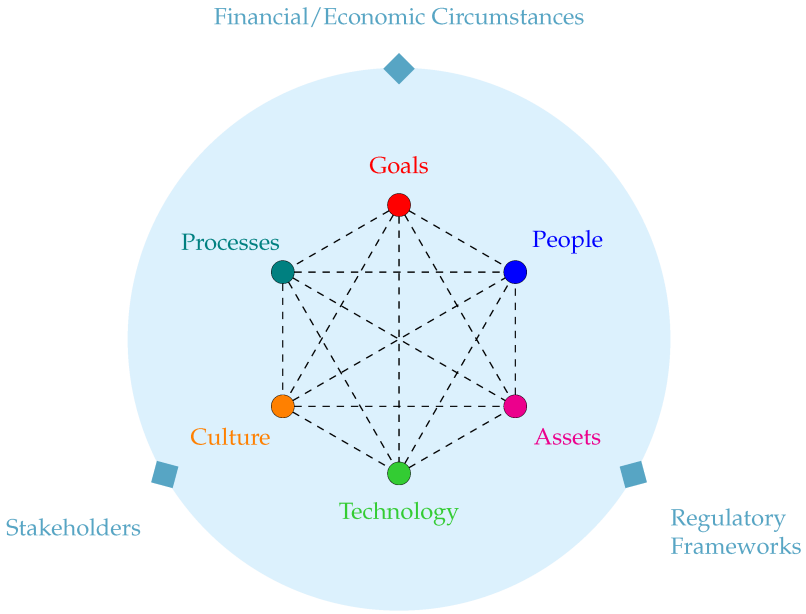


Figure 3.3: A socio-technical system embedded within an external environment, based on Davis et al. (2014).

more mature SME categories to have achieved the goals of less mature SME categories.

Our framework was constructed based on earlier cybersecurity frameworks focusing on SMEs (Benz and Chatterjee, 2020; Carías, Borges, et al., 2020; Cholez and Girard, 2014) or STS (AlHogail, 2015; Da Veiga et al., 2020; Malatji, Marnewick, et al., 2020; Malatji, Von Solms, et al., 2019; Sittig and H. Singh, 2016). Interestingly, none of these frameworks focused on both SMEs and STS. To address the singular characteristics of our setting, we additionally incorporated the findings from our systematic review, as well as principles for designing cybersecurity maturity models for SMEs (Yigit Ozkan and Spruit, 2020), in our framework. Our findings appear most prominently in the 'Technology' aspect, explaining why this column of Table 3.10 contains relatively few references to earlier work.

Our results relating to the various ADKAR dimensions serve as input for the 'People' and 'Culture' aspects. Start-ups and digitally dependent SMEs should focus on making their employees aware and providing initial cybersecurity knowledge to inspire desire and motivation. This can be achieved through a culture of organisational commitment to cybersecurity (AlHogail, 2015; Da Veiga et al., 2020). Digitally based SMEs and digital enablers should progress through the ADKAR phases, with the aid of cybersecurity training, policy, and assessment. Eventually, employees should mutually reinforce each

Table 3.10: Socio-technical cybersecurity framework for SMEs.

| | Socio-Technical Aspects | | | | | |
|---|---|---|---|---|---|---|
| SME CATEGORY | GOALS | PEOPLE | CULTURE | PROCESSES | TECHNOLOGY | ASSETS |
| **Start-ups** | Realise cybersecurity necessity (Malatji, Von Solms, et al., 2019) due to external environment factors. Move from a non-existent cybersecurity culture to initial, informal cybersecurity measures (Benz and Chatterjee, 2020; Malatji, Marnewick, et al., 2020; Malatji, Von Solms, et al., 2019). | Define training plans and start creating cybersecurity awareness (Carías, Borges, et al., 2020). | Initial cybersecurity policies and procedures show management commitment, ensuring employee support (AlHogail, 2015; Da Veiga et al., 2020). | No standardised processes yet (Malatji, Von Solms, et al., 2019). SME gains awareness on cybersecurity policies, processes, procedures, standards and regulation. | Employ a threat-based risk assessment tool requiring no knowledge of SME assets, using no/intuitive aggregation. External support to understand and implement countermeasures. | Understand relevant and critical cybersecurity asset types (Carías, Borges, et al., 2020). |
| **Digitally dependent** | Start formalising cybersecurity processes. Define, manage, and communicate cybersecurity strategy (Benz and Chatterjee, 2020; Carías, Borges, et al., 2020; Malatji, Marnewick, et al., 2020; Malatji, Von Solms, et al., 2019). | Continue building awareness (Da Veiga et al., 2020). Stimulate desire through knowledge acquisition (AlHogail, 2015). Evaluate gaps in ability (Carías, Borges, et al., 2020). | Management support and cybersecurity trainings stimulate employees (Da Veiga et al., 2020) and change their perception (AlHogail, 2015). | Formulate basic (reactive) cybersecurity policies, processes, and procedures (Da Veiga et al., 2020; Malatji, Von Solms, et al., 2019); likely not yet universally applied across business units (Malatji, Von Solms, et al., 2019). | Employ a threat-based risk assessment tool using no/intuitive aggregation. External support to implement countermeasures. | Systematically identify and document relevant assets and their baseline configurations (Carías, Borges, et al., 2020). |
| **Digitally based** | Establish a formal cybersecurity programme that facilitates continuous improvement and compliance with regulation (Benz and Chatterjee, 2020; Carías, Borges, et al., 2020; Malatji, Marnewick, et al., 2020; Malatji, Von Solms, et al., 2019). | Advance cybersecurity knowledge and ability through clearly communicated and documented trainings (Carías, Borges, et al., 2020; Da Veiga et al., 2020; Malatji, Von Solms, et al., 2019). | Regular communication and education (Da Veiga et al., 2020), backed by rewards and deterrents (AlHogail, 2015), ensures secure employee behaviour (AlHogail, 2015; Da Veiga et al., 2020). | Processes defined and documented proactively, communicated via awareness and training sessions (Da Veiga et al., 2020; Malatji, Von Solms, et al., 2019). Information sharing agreements defined (Carías, Borges, et al., 2020). | Use a risk assessment framework or maturity model with adequately motivated aggregation. Implement basic countermeasures (Carías, Borges, et al., 2020), external support for complex countermeasures. | Manage asset changes and periodically maintain assets (Carías, Borges, et al., 2020). |
| **Digital enablers** | Embed and automate cybersecurity processes (Benz and Chatterjee, 2020; Cholez and Girard, 2014; Malatji, Marnewick, et al., 2020; Malatji, Von Solms, et al., 2019), which, combined with collaborative stakeholder relationships (Carías, Borges, et al., 2020), promote internal and external trust in the SME cybersecurity posture (Da Veiga et al., 2020). | Employees mutually reinforce their cybersecurity abilities, possibly captured in official cybersecurity roles (Da Veiga et al., 2020). | Regular evaluations (Cholez and Girard, 2014; Malatji, Von Solms, et al., 2019) stimulate naturally secure behaviour (Da Veiga et al., 2020), where national culture and regulations are recognised (AlHogail, 2015). An environment of trust with stakeholders exists (Carías, Borges, et al., 2020; Da Veiga et al., 2020). | Successive comparisons of assessment results facilitate continuous process improvement (Cholez and Girard, 2014; Malatji, Von Solms, et al., 2019). Business continuity plan defined and communicated to external stakeholders (Carías, Borges, et al., 2020). | Use a risk assessment framework or maturity model with advanced aggregation. Independently implement countermeasures (Carías, Borges, et al., 2020) and actively detect anomalies (Carías, Borges, et al., 2020), with the help of automated tools (Malatji, Von Solms, et al., 2019). | Identify and document internal and external dependencies of assets, to help in determining the SME attack surface. Actively monitor assets (Carías, Borges, et al., 2020). |

other's cybersecurity abilities (Da Veiga et al., 2020). The ideal cybersecurity culture will lead to trust from both the people inside the SME, as well as the environment outside of the SME (Carías, Borges, et al., 2020; Da Veiga et al., 2020).

Start-ups and digitally dependent SMEs are often not aware of the existence of cybersecurity standards (European DIGITAL SME Alliance, 2020). These SMEs should first become aware and then begin to formulate basic cybersecurity policies, processes, and procedures (Da Veiga et al., 2020; Malatji, Von Solms, et al., 2019). Digitally based SMEs should have formal processes in place to reinforce desired cybersecurity behaviour of employees (Malatji, Von Solms, et al., 2019). Digital enabler SMEs should strive towards continuous process improvement (Cholez and Girard, 2014; Malatji, Von Solms, et al., 2019), which enables business continuity (Carías, Borges, et al., 2020).

We map the 'Technology' aspect of STS to the advised cybersecurity assessment approach and tooling for the SME. This is in line with the approach of Malatji, Von Solms, et al. (2019), who incorporate "cybersecurity tools and resources" in the 'Technology' aspect of their socio-technical cybersecurity framework.

Start-ups should understand relevant cybersecurity asset types and digitally dependent SMEs should begin identifying and documenting assets (Carías, Borges, et al., 2020). Without an asset inventory or internal cybersecurity expertise, most risk assessment and maturity model approaches are not suited to these SMEs. Additionally, they are just beginning to cultivate a desire among employees to improve cybersecurity. Incorporating the real-life threat environment (Gollmann et al., 2015) is an attractive option to promote motivation. Focusing on the real-life threat environment can increase the feelings of task relevance and significance employees feel, which are key motivators (Kam et al., 2020). This is why we advise a threat-based cybersecurity risk assessment approach for start-ups and digitally dependent SMEs.

In the same vein, we advise to not aggregate scores in cybersecurity assessment solutions for start-ups and digitally dependent SMEs. If aggregation is deemed necessary, injective and idempotent aggregation strategies should be used, such as WLC and WM. Strategies that satisfy injectivity and idempotence can be seen as intuitive. Using these strategies allows for feelings of competence and relatedness among employees, which stimulate motivation (Menard et al., 2017). This puts employees in a position to be a part of the solution to SME cybersecurity challenges, rather than being the source of the challenges (Zimmermann and Renaud, 2019).

The combination of simple aggregation and a threat-based approach offers another benefit: the corresponding assessments do not necessarily require extensive internal expertise and data. Many of the more complex aggregation strategies and comprehensive assessment approaches require cybersecurity experts at the SME to determine parameters and weights. Such resources are

limited at SMEs (Heidt et al., 2019), and especially at start-ups and digitally dependent SMEs. This is why assessment approaches for these SMEs should preferably be largely based on data that can be automatically collected. Threat-based approaches are ideally suited to this requirement, as general incident data is widely available (Y. Liu et al., 2015), and can be mapped to threats to offer SMEs insight into what is important for them (Casola et al., 2019).

Digitally based SMEs and digital enablers can be expected to have a complete inventory of assets (Carías, Borges, et al., 2020). Digital enablers should additionally be aware of internal and external dependencies (Carías, Borges, et al., 2020), allowing them to specify their attack surface (Manadhata and Wing, 2011). For these SME categories, complete risk- and maturity assessments are desirable. Digital enablers will often require comprehensive assessments that can prove compliance with cybersecurity standards and regulations.

Digitally based SMEs should consider using aggregation strategies that reflect desirable security properties, such as the weakest link principle. Using a WCP strategy can guide these SMEs towards more accurate assessments, although intuitiveness is sacrificed. Digital enablers with cybersecurity expertise, a specified attack surface, and large volumes of internal data, should consider more advanced aggregation strategies.

Figure 3.4 provides a visual summary of the STS interactions inherent to our framework. We use coloured arrows to indicate interactions that are explicitly mentioned in Table 3.10. It is implicit in the STS model of Davis et al. (2014) that all aspects are interrelated.

The direction of the arrows indicates which aspect serves as an input for another aspect. For start-ups, the external environment aspects motivate the SME to realise the necessity of investing in cybersecurity, leading to the initial goals. For digitally dependent SMEs, the goals formulated by management serve as catalysts for culture and processes. We observe that from an initial external motivation for start-ups, SMEs gradually build up internal interactions. For digital enablers, we see many interactions, both internally and with the external environment.

## 3.6 DISCUSSION

We extensively analysed and interpreted our results in Sections 3.4 and 3.5. This section will focus on a discussion of our research questions and the potential limitations of our research.

Our first research question asked: how are cybersecurity metrics aggregated in socio-technical cybersecurity measurement solutions? One interesting finding from Table 3.8 is that half of the research involving implementations did not aggregate at all. Table 3.2 gives a partial explanation for this phenomenon: no aggregation strategy satisfies all desirable security properties. Thus, aggregation should preferably be avoided. Nevertheless, aggregation using basic approaches such as WLC is prevalent, with 42 of our 60 inclusions

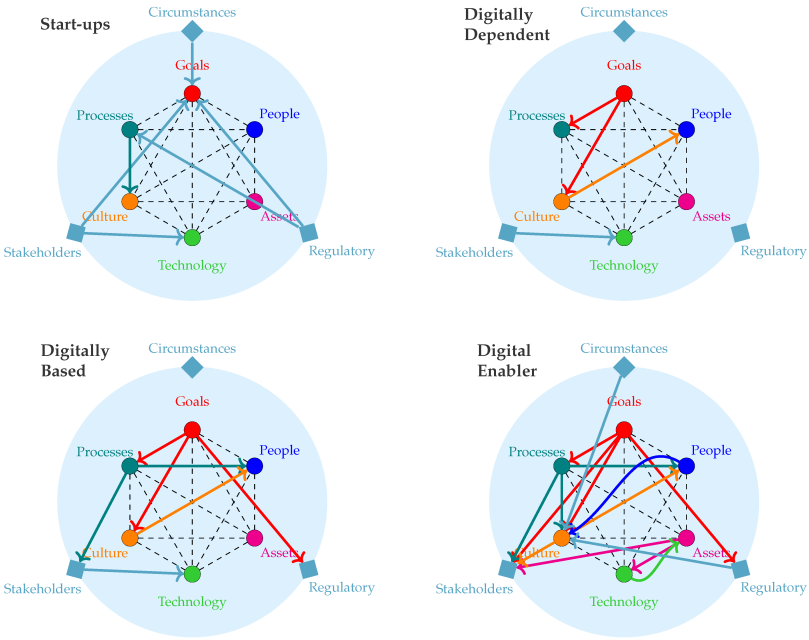


Figure 3.4: A visualisation of the framework presented in Table 3.10 using the representation of Figure 3.3.

using this aggregation technique. We observed a clear lack of dependency consideration among metrics, which could be solved using Bayesian network (Dantu and Kolan, 2005; Dantu, Kolan, and Cangussu, 2009; N. Feng et al., 2014; Sahinoglu, 2008) or ANP techniques (Brožová et al., 2016; Lo and W.-J. Chen, 2012). Our cybersecurity framework presented in Table 3.10 provides clear guidance on which aggregation strategies suit which SME categories.

Our second research question was formulated as: How do aggregation strategies differ in cybersecurity measurement solutions relevant to SMEs and all other solutions? Our analysis of Table 3.9 demonstrated that in enterprise research little to no attention is paid to aggregation strategies that satisfy the weakest link and dependency properties. One of the main obstacles in making aggregation strategies suitable for SMEs is the time and expertise required to carry them out. Generally, more complex aggregation strategies require the determination of more parameters and relationships, which in turn often requires consultation of security experts at the cyber-system being assessed (for example, (Alencar Rigon et al., 2014; Damenu and Beaumont, 2017; Proença and Borbinha, 2018; Shokouhyar et al., 2018)). This expertise is rarely available at smaller SMEs, although when it is, ANP approaches (Brožová et al., 2016; Lo and W.-J. Chen, 2012) could offer a path towards more accurate aggregation.

Our final research question covered the consideration of adaptability: "the state of being able to change to work or fit better" (J.-H. Cho et al., 2019). We found that very few papers consider the effects of missing data, dirty data, security events, or concept drift; all are vital elements in determining the ability of a solution to adapt to unexpected circumstances to work better. Research does often recognise the need for being able to change to fit better, as shown by the relatively large proportion that considers adaptation to other use cases. Nevertheless, there is still much to be gained in this area. It is vital that authors of research on socio-technical cybersecurity measurement solutions explicitly address the adaptability dimension in the future. Our framework of Table 3.10 helps in this regard, with its focus on proactive processes and active monitoring and detection capabilities.

We additionally analysed the ADKAR factors that were addressed in our inclusions. We found that desire was rarely considered in research. This was especially true for research focusing on the defender perspective. Additionally, we found that the real-life threat environment, as defined in Gollmann et al. (2015), is considered in less than half of our inclusions. Both of these findings offer an interesting contrast to the increasingly important role SDT and PMT play in security research (Menard et al., 2017). These theories focus heavily on (intrinsic) motivation and threat perception (Padayachee, 2012). Given the low intrinsic motivation among SMEs and their employees to improve security (Heidt et al., 2019), and the relatively large impact individual employees can have in the SME context, future research focusing on motivation and the real-life threat environment could provide an interesting avenue for making cybersecurity solutions more suitable to SMEs.

### 3.6.1 *Limitations and threats to validity*

We should mention at this stage that our research is not without its limitations. One potential issue is that our systematic review was not restricted to recent years, which meant that contemporary research was not as prominent in this review as it is in most other reviews. This could mean that we are overlooking certain recent developments, although 18 of our 60 inclusions were published in the past three years.

Additionally, although we believe our 60 inclusions are sufficient to help us answer our research questions, certain groupings of the inclusions resulted in relatively small sub-samples from which to draw conclusions. This could limit the generalisability of our analysis and conclusions, meaning that one could have different findings when considering different cybersecurity focus areas.

We believe in the construct validity of our systematic review methodology SYMBALS (van Haastrecht, Sarhan, Yigit Ozkan, et al., 2021), as it is based on widely-accepted methods (van de Schoot et al., 2021; Wohlin, 2014) and guidelines (Kitchenham and Charters, 2007; Liberati et al., 2009; Moher et al.,

2015). However, it is still a novel methodology that remains to be extensively tested. We feel this does not threaten the validity of our research, since SYMBALS is geared towards reproducibility and satisfies standard reporting item guidelines for systematic reviews (Moher et al., 2015).

A final mention should be made of our choice to approach the social dimension through the ADKAR change management model (Hiatt, 2006). Although the model has been applied in the cybersecurity domain (Da Veiga, 2018), it is certainly not a standard approach to use ADKAR in this setting. Nevertheless, Table 3.5 summarised the natural mapping of social cybersecurity metric concepts to the ADKAR framework and our framework presented in Table 3.10 showed how the ADKAR terms can be instinctively imported from previous research. Hence, we feel justified in using this approach.

## 3.7 CONCLUSION AND FUTURE RESEARCH

Businesses, and especially small- and medium-sized enterprises (SMEs), struggle to cope with the existing cyber threat landscape. Researchers have turned to cybersecurity measurement to deal with these issues, although many challenges remain, such as how to aggregate sub-metrics into higher-level metrics (J.-H. Cho et al., 2019). The challenges faced by SMEs are compounded by the dynamic nature of the cyber threat landscape, necessitating adaptable solutions. These current challenges motivated us to investigate the topics of aggregation and adaptability in this review, with a focus on SMEs.

The social side of cybersecurity deserves attention, certainly in the SME context. This is why we chose to direct our review at socio-technical cybersecurity measurement solutions. The ADKAR (Awareness, Desire, Knowledge, Ability, Reinforcement) change management model of Hiatt (2006) guided us in covering the social dimensions considered in research. To aid in the analysis of aggregation approaches, we outlined five main aggregation strategy classes in Section 3.2.3: weighted linear combinations, weighted products, weighted maxima, weighted complementary products, and Bayesian networks. We looked towards existing research to determine interesting dimensions of adaptability, such as missing or dirty data (W. Kim et al., 2003) and concept drift (Widmer and Kubat, 1996).

Based on our analysis in Sections 3.2.3 and 3.4, we found that aggregation should only be carried out if necessary, since no single aggregation strategy exists that satisfies all of the desired security properties. Notably, dependencies among metrics are often not considered. Solutions can be found in this area in Bayesian networks (Dantu and Kolan, 2005; Dantu, Kolan, and Cangussu, 2009; N. Feng et al., 2014; Sahinoglu, 2008) and analytic network process (Brožová et al., 2016; Lo and W.-J. Chen, 2012) techniques.

We used our findings as input to construct a socio-technical cybersecurity framework for SMEs. We presented our framework in Table 3.10 and visualised it in Figure 3.4. Offering a single solution for all SMEs is too simplistic. This

is why we divided SMEs into four categories, as suggested by the European DIGITAL SME Alliance (2020): start-ups, digitally dependent SMEs, digitally based SMEs, and digital enablers. By detailing what can be expected of each SME category, we were able to determine which cybersecurity assessment strategies were suitable in each case. For start-ups and digitally dependent SMEs, threat-based risk assessment approaches that either do not aggregate or use intuitive aggregation strategies are ideal. By focusing on the real-life threat environment (Gollmann et al., 2015), relevance and significance of the assessment task are given a central role. A simple and intuitive aggregation strategy accommodates feelings of competence and relatedness. Altogether, this ensures optimal organisation and employee motivation (Kam et al., 2020; Menard et al., 2017).

Digitally based SMEs and digital enablers are advised to use more comprehensive risk assessment approaches and maturity models. These assessment techniques should assist in working towards or proving compliance with standards and regulations. Under ideal circumstances, this will build trust in the cybersecurity posture of the SME, both internally and externally. Digital enablers are also prime candidates for using more advanced aggregation strategies such as Bayesian networks, since they often have the cybersecurity expertise and data required to make these solutions successful.

We hope that our socio-technical cybersecurity framework will provide a basis to design successful cybersecurity assessment solutions for SMEs. SMEs should not be forced to use solutions that are not suited to their situation. Especially start-ups and digitally dependent SMEs currently lack suitable cybersecurity assessment solutions, even though they are most in need of "easily understandable and practical solutions" (European DIGITAL SME Alliance, 2020). In future work, we aim to help these SMEs to become more secure. An important first step is to formulate a properly motivated, intuitive, and usable threat-based cybersecurity risk assessment approach, to offer this most vulnerable group some deserved cybersecurity respite.